

Shared genetics between breast cancer and its predisposing diseases identify novel breast cancer treatment candidates

Panagiotis N. Lalagkas & Rachel D. Melamed

Department of Biological Sciences, University of Massachusetts, Lowell, MA, USA

Corresponding author: Rachel D. Melamed, Rachel.Melamed@uml.edu

Abstract

Breast cancer is a major public health issue. Current treatment options, while effective, have severe side effects. FDA-approved drugs have known safety and pharmacological profiles and some, such as metformin, have been tested in clinical trials for repurposing for breast cancer. However, clinical trials are slow and expensive, which creates the need for innovative approaches to accelerate drug repurposing for breast cancer. We have previously shown that genes associated with Mendelian diseases that predispose patients to certain cancers are enriched for successful drug targets, due to their pleiotropic effects. Here, we extend our approach to exploit clinical associations between breast cancer and its predisposing diseases for drug repurposing. We hypothesize that pleiotropic genes shared between breast cancer and its predisposing diseases can help us discover new uses for drugs currently approved only for the predisposing diseases. To test our hypothesis, we compile a list of six traits known to increase breast cancer risk (predisposing diseases). Using GWAS summary statistics and local genetic correlation analysis, we find 84 genomic loci harboring mutations with positively correlated effects between breast cancer and each predisposing disease. These loci contain 194 protein-coding genes (shared genes). Using a network biology approach and canonical pathways, for each disease pair, we connect drugs already indicated for the predisposing disease to its shared biology with breast cancer and identify drug repurposing candidates for breast cancer. Finally, we show that our list of candidate drugs is enriched for currently investigated and indicated drugs for breast cancer (OR=9.28, $p=7.99e-03$). Our findings suggest a novel way to accelerate drug repurposing for complex diseases by leveraging shared genetics.

Keywords

breast cancer, shared genetics, predisposing disease, risk factors, drug repurposing, drug genetics

Introduction

Breast cancer is a major public health issue and the most common type of cancer in women worldwide¹. Current treatment options, such as chemotherapy, hormone therapy, and immunotherapy², while effective, often have severe side-effects that impair patients' quality of life and adherence to treatment^{3,4}. On the other hand, existing FDA-approved drugs have established safety and pharmacological profiles, and testing them for repurposing for breast cancer treatment has been an attractive strategy. For instance, metformin, the most

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

commonly prescribed anti-diabetic drug, has shown anti-proliferative effects in pre-clinical studies and is currently undergoing clinical trials for breast cancer treatment. This surprising effect may be explained by its downstream effects activating *AMPK* protein kinase, a tumor suppressor critical in regulating cell proliferation^{6,7}. However, oncology clinical trials are slow, expensive and have up to 95% attrition rates, mainly due to lack of understanding of the mechanism of action of the tested drug prior to the clinical trial⁸⁻¹⁰. Therefore, new approaches are needed to both accelerate drug repurposing for breast cancer treatment and provide biological insights to design clinical trials with higher success rates.

Genome-wide association studies (GWAS) associate genes with a disease; drugs targeting those genes are twice as likely to successfully treat that disease¹¹⁻¹³. For example, a recent breast cancer GWAS found a strong signal in the *ESR1* locus which is the target of tamoxifen, a widely used drug for breast cancer treatment¹⁴. This highlights the great potential of genetics in informing drug discovery and repurposing¹⁵. However, the inherent limitations of GWAS, such as the need for large sample sizes and the challenge of identifying true causal genes, limit their ability to inform drug repurposing^{16,17}. This creates the need for innovative ways to build on genetics findings to discover drugs for diseases of high public importance, such as breast cancer.

To this end, we propose that *shared genetics between breast cancer and other diseases* can inform drug repurposing for breast cancer. This idea is supported by prior research. For one, it has been shown that clinically co-occurring diseases share genetics^{18,19}. This suggests that shared biological processes may explain why a health condition might predispose individuals to breast cancer. Consequently, drugs approved for such predisposing diseases and targeting shared biology with breast cancer might hold potential for breast cancer treatment. In further support of this concept, we have previously built on work showing bearers of Mendelian disease mutations may suffer increased risk of certain cancers, due to a hidden role of the Mendelian disease genes in the comorbid cancer. In a follow-up study, we showed that the Mendelian disease genes implicated for a cancer are enriched for successful drug targets²⁰. Building on these observations, here, we propose the repurposing of drugs based on shared genetic etiology.

Ultimately, we propose 59 drugs that have not been previously tested for breast cancer, accompanied by biological insights supporting their potential therapeutic effect on breast cancer.

Results

Overview of approach

Health conditions including high cholesterol and type 2 diabetes incur increased risk for breast cancer. Previous genetics research has supported a biological explanation: these diseases, which we will refer to as *predisposing diseases*, share genetic variation with breast cancer. Building on this, we hypothesize that finding the shared genetics between breast cancer and its predisposing diseases can help us discover new drugs for breast cancer. Specifically, we hypothesize that drugs approved for a predisposing disease and targeting its shared biology with breast cancer can treat the latter disease (**Figure 1A**).

To test this hypothesis, we first search the scientific literature (epidemiological and statistical genetic studies) for diseases with genetic variation known to predispose individuals to breast cancer. We find six such diseases (**Table 1**).

Next, we aim to identify the shared genetics between each pair of breast cancer and a predisposing disease that can be tied to medications currently used to treat the predisposing disease. We use publicly available GWAS summary statistics data (**Table 1**) and a local genetic correlation analysis, to find the shared genetics. Following the identification of likely shared genes for each disease pair, we connect them to drugs currently used to treat the predisposing disease. Because we wish to both make this connection and gain interpretation about the pathways involved, we link each shared gene, and each set of drug targets, to canonical pathways using a network biology approach. By finding drugs that target shared pathways, we both prioritize candidate drugs for repurposing for breast cancer and provide biological insights that support their effect in disease treatment. **Figure 1B** illustrates the complete workflow using the example of high HDL and breast cancer.

Finally, we evaluate our list of candidate drugs. To do so, we compile a list of 583 drugs either in clinical trials (N=451) or approved (N=132) for breast cancer and test for enrichment within our candidate drugs.

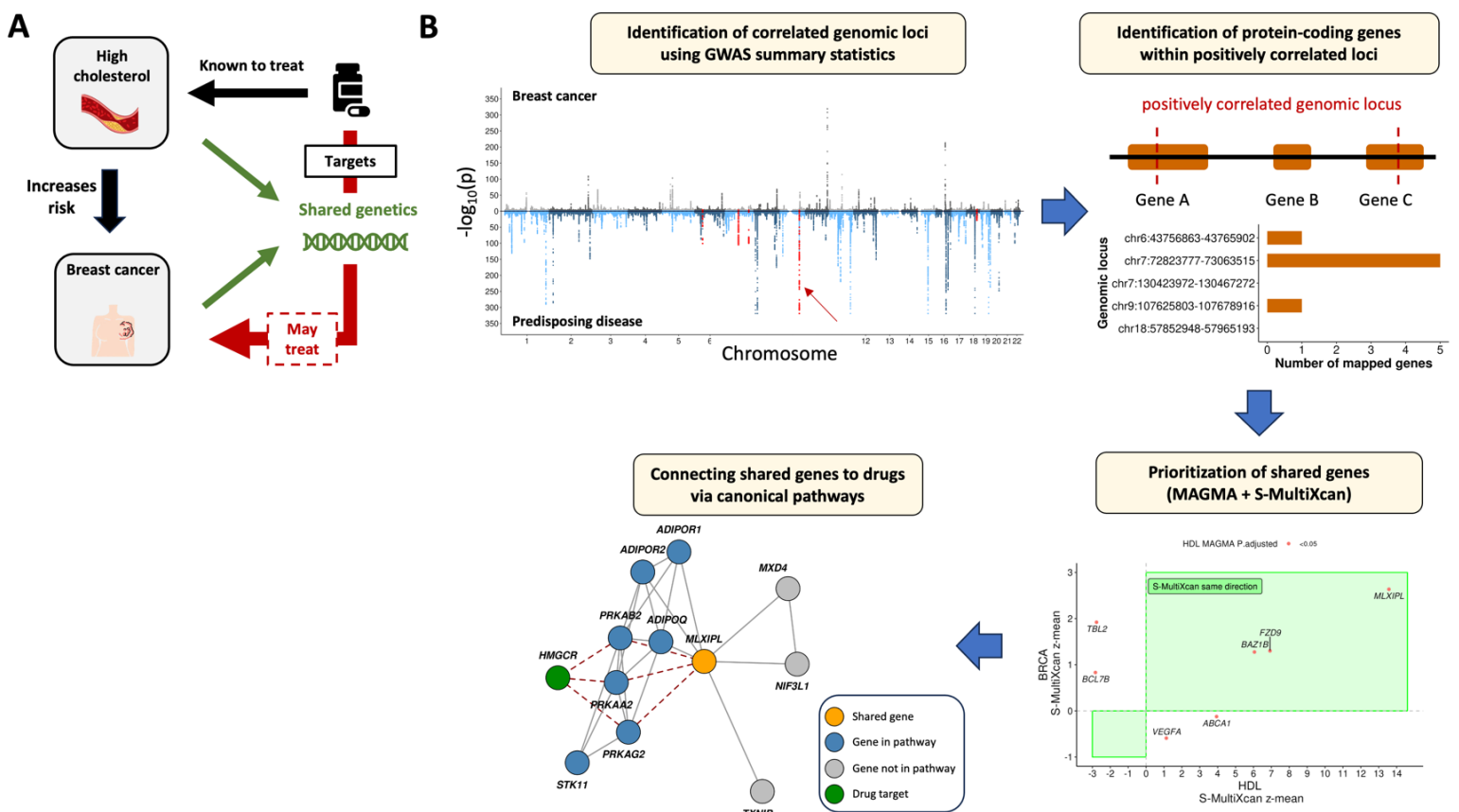


Figure 1. Outline of the approach. **A.** Tested hypothesis that a drug treating a health condition known to increase the risk for breast cancer and targets the shared biology of both conditions may also treat breast cancer. **B.** Proposed workflow to identify the most likely shared genes between breast cancer and a predisposing disease and connect them to drugs approved for the predisposing disease.

Disease	Consortium of GWAS	Year of publication	Ncases	Ncontrols	Ntotal	Supported connection with breast cancer
Breast cancer	BCAC	2020	133,384	113,789	247,173	-
Depression	PGC	2019	170,756	329,443	500,199	Zhu et al. ²¹
High density cholesterol	GLGC	2021	-	-	1,209,154	Johnson et al. ²²
Low density cholesterol	GLGC	2021	-	-	1,199,124	Johnson et al. ²²
Prostate cancer	PRACTICAL	2018	79,194	61,112	140,306	Ren et al. ²³
Schizophrenia	PGC	2022	53,386	77,258	130,644	Byrne et al. ²⁴
Type 2 diabetes	DIAMANTE	2022	80,154	853,816	933,970	Pearson-Stuttard et al. ²⁵

Table 1. Evidence for shared etiology of predisposing conditions, and genome-wide association study (GWAS) summary statistics for breast cancer and its six predisposing diseases used in this study. All GWAS include samples of European ancestry.

BCAC: Breast Cancer Association Consortium; PGC: Psychiatric Genomics Consortium; UKBB: United Kingdom BioBank; GLGC: Global Lipids Genetics Consortium; PRACTICAL: Prostate Cancer Association Group to Investigate Cancer-Associated Alterations in the Genome; DIAMANTE: Diabetes Meta-Analysis of Trans-Ethnic association studies

Discovering shared genetics between breast cancer and its predisposing diseases

The first step in testing our hypothesis is to identify genomic loci likely to drive the shared etiology between breast cancer and a predisposing disease. Although cross-trait Linkage Disequilibrium Score Regression (LDSC) is a widely adopted method to identify genetic correlation between a pair of phenotypes using GWAS summary statistics, its genome wide nature means it does not provide insight into particular genes driving the correlation. To gain that insight, we perform a local genetic correlation analysis for each pair of breast cancer and predisposing disease, using LOGODetect²⁶. Across all disease pairs, we identify 59 negatively (per disease pair: median=5.5; min=0; max=34) and 84 positively (per disease pair: median=10; min=5; max=37) correlated genomic loci (**Figure 2A, Supplementary Tables 1 & 2**). Notably, the identified loci are distributed across all autosomal chromosomes and not localized in specific parts of the genome (**Figure 2B**). In order to confirm that LOGODetect does not discover shared loci for disease pairs without known epidemiological or clinical associations, such as high LDL-depression and high LDL-schizophrenia, we repeat the analysis for these pairs. LOGODetect does not identify any significant correlated loci.

Next, for each predisposing disease, we seek to prioritize genes in the correlated genomic loci that are the most likely drivers of shared etiology with breast cancer. Since our ultimate goal is to recommend candidate drugs for repurposing for breast cancer, we are more

interested in genes exhibiting effects in the same direction in both breast cancer and predisposing disease. Therefore, for the downstream analysis, we focus on genomic loci found to be positively correlated for each disease pair. Using the SNP2GENE function from FUMA²⁷, we extract a total of 194 protein-coding genes that fall within all the positively correlated loci (median of 37 protein-coding genes per disease pair; min=8; max=81). For each disease pair, these genes constitute the list of shared genes that likely drive the shared etiology with breast cancer.

From this list, we seek the subset most likely to contribute to druggable shared etiology. That is, we wish to prioritize genes more likely to 1) participate in shared pathophysiological processes and 2) relate to the effect of drugs on the predisposing disease. First, we use MAGMA, a tool that aggregates the effect of all SNPs within a gene, to keep genes significantly associated with the predisposing disease in every disease pair²⁸. This filter is important as the predisposing disease increases the risk for breast cancer, implying that any genes underlying that shared risk should, at a minimum, impact the predisposing disease. Second, we use S-MultiXcan, a tool that uses GWAS summary statistics data to genetically predict the expression of a gene summarized across 49 GTEx tissues²⁹. We keep only genes with genetically predicted expression consistently aligned in the same direction (either downregulated or upregulated) in both diseases in a disease pair, as dysregulation in opposite directions is not easily interpretable and not useful for therapeutic purposes. For instance, *ABCA1* is found to be one of the genes shared between high LDL and breast cancer. Probuocol, an approved anti-cholesterol drug, inhibits *ABCA1*. But, *ABCA1* is upregulated in high LDL and downregulated in breast cancer, suggesting that inhibitions of this gene might not lead to desired treatment outcome for breast cancer³⁰. **Figure 2C** shows the number of shared genes within positively correlated loci between breast cancer and each predisposing disease, before and after applying the filters. The full list of shared genes for each disease pair is provided in **Supplementary Table 3**.



Figure 2. Shared genetics between breast cancer and its predisposing diseases. **A.** Number of negatively and positively correlated loci between each breast cancer and predisposing disease pair. **B.** Position in the genome of the identified correlated loci for each breast cancer and predisposing disease pair. The y-axis represents the $-\log_{10}(\text{q-value})$ of each locus as provided by LOGODetect. **C.** Number of protein-coding genes located within the positively correlated loci for each breast cancer and predisposing disease pair, before and after applying the MAGMA & S-MultiXcan filters.

Connecting shared genetics to drugs

The next step is to connect the identified shared genes to drugs treating the predisposing diseases. We make this connection through canonical pathways. In that way, we can capture the biological processes which 1) are involved with identified shared genes and 2) are related to drugs indicated for the predisposing disease. First, for each breast cancer and predisposing disease pair, we connect the identified shared genes to canonical pathways using the STRING protein-protein interaction network and a network propagation algorithm (**Supplementary Table 4**). Notably, our workflow identifies pathways with known roles in tumorigenesis and progression that are also disrupted in the breast cancer predisposing diseases (**Figure S1**). For instance, we identify the *WNT signaling pathway* to be part of the shared etiology between breast cancer³¹ and each of high HDL³², prostate cancer³³ and type 2 diabetes³⁴. Additionally, the *mTOR signaling pathway*, another cancer-related pathway^{35,36}, is found to be shared between breast cancer and high LDL³⁷.

After connecting the shared genes for every disease pair to pathological processes, we seek to find which drugs target them. To do so, we use a similar approach and connect

targets of a drug currently approved for a predisposing disease to the shared canonical pathways between breast cancer and that disease (**Figure 3, Supplementary Table 5**). Drugs significantly connected to at least one shared canonical pathway are considered candidate drugs for repurposing for breast cancer.

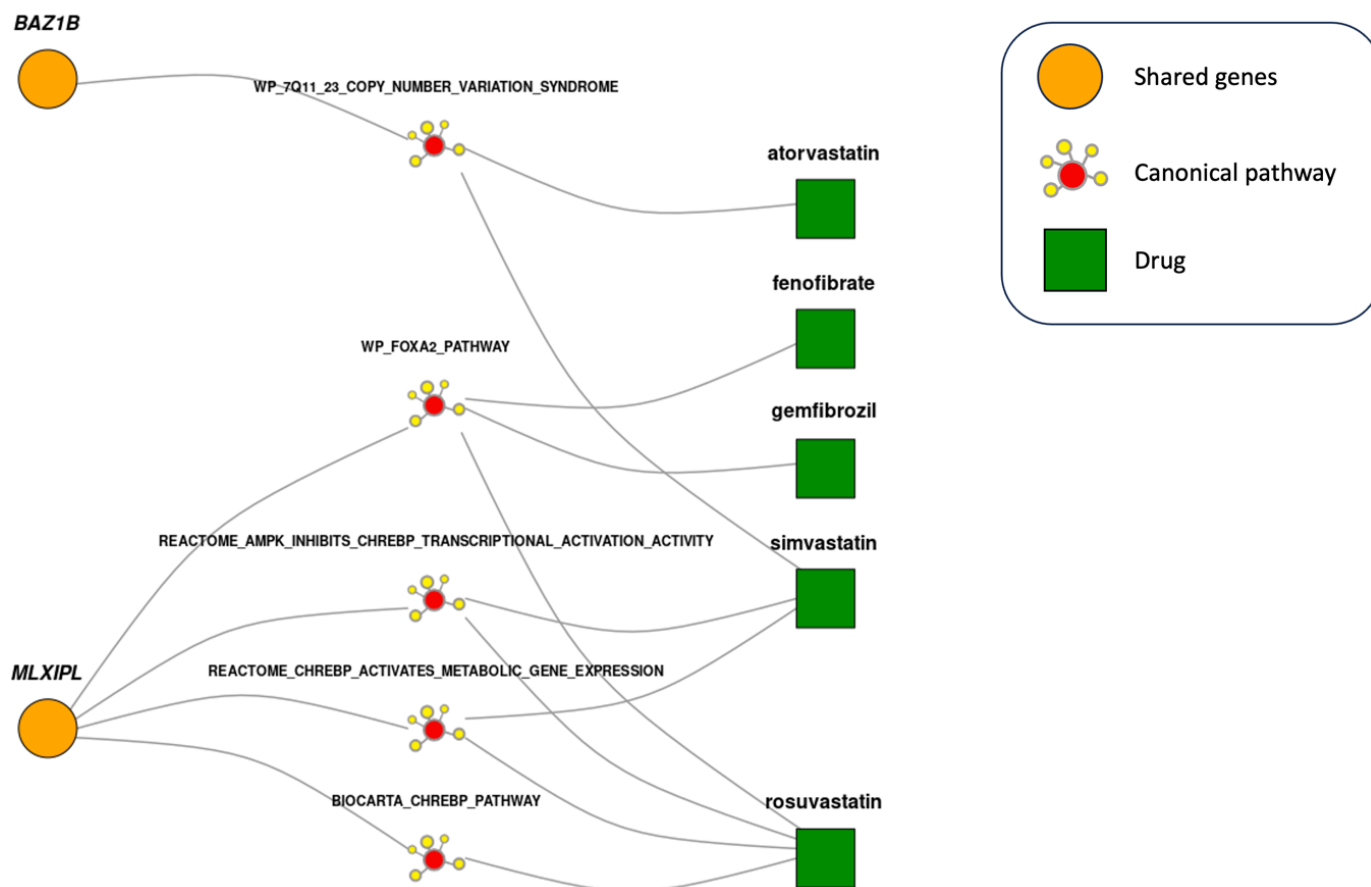


Figure 3. Drugs treating high HDL and targeting its shared biology with breast cancer. These drugs are recommended for repurposing for breast cancer and the shared pathways they target provide a biological basis for the repurposing. Only canonical pathways significantly linked to both identified shared genes and high HDL drugs are shown.

Evaluation of drug repurposing and prioritization of new candidates

The final step is to assess the efficacy of our approach in identifying promising candidate drugs for repurposing for breast cancer. To do so, we compare our list of candidate drugs to those currently under investigation or approved for breast cancer treatment. In total, out of 112 approved drugs for the six breast cancer predisposing diseases, 16 have undergone testing or received approval for breast cancer treatment (**Figure 4A**). Remarkably, our method identifies 15 out of these 16 drugs, while also offering insights into the specific genes and pathways that might explain their effect on breast cancer (**Supplementary Table 5**). When considering all the recommended candidate drugs for repurposing, we find a significant enrichment for drugs currently investigated or approved for breast cancer (OR=9.28, $p=7.99e-03$, one-sided Fisher's exact test).

Our workflow identifies *HMGCR* to be shared between high LDL and breast cancer. However, *HMGCR* is also the target of statins, a group of drugs approved for lowering the LDL blood levels. Therefore, to ensure that the significance of our results is not driven by this instance where a shared gene is also the target of an approved drug for the predisposing disease, we repeat the analysis by excluding the high LDL-breast cancer pair. Again, we find strong enrichment (OR=9.35, $p=9.66e-03$, one-sided Fisher's exact test), highlighting the ability of our approach in connecting shared biology to drugs even when a shared gene is not directly targeted by a drug.

In conclusion, we recommend 76 candidate drugs for repurposing while providing biological insights (specific genes and pathways) supporting their potential use in breast cancer treatment. Among these candidates, 15 are already in advanced clinical trial phases or approved for breast cancer treatment, while 59 are novel candidate drugs (**Figure 4B**).

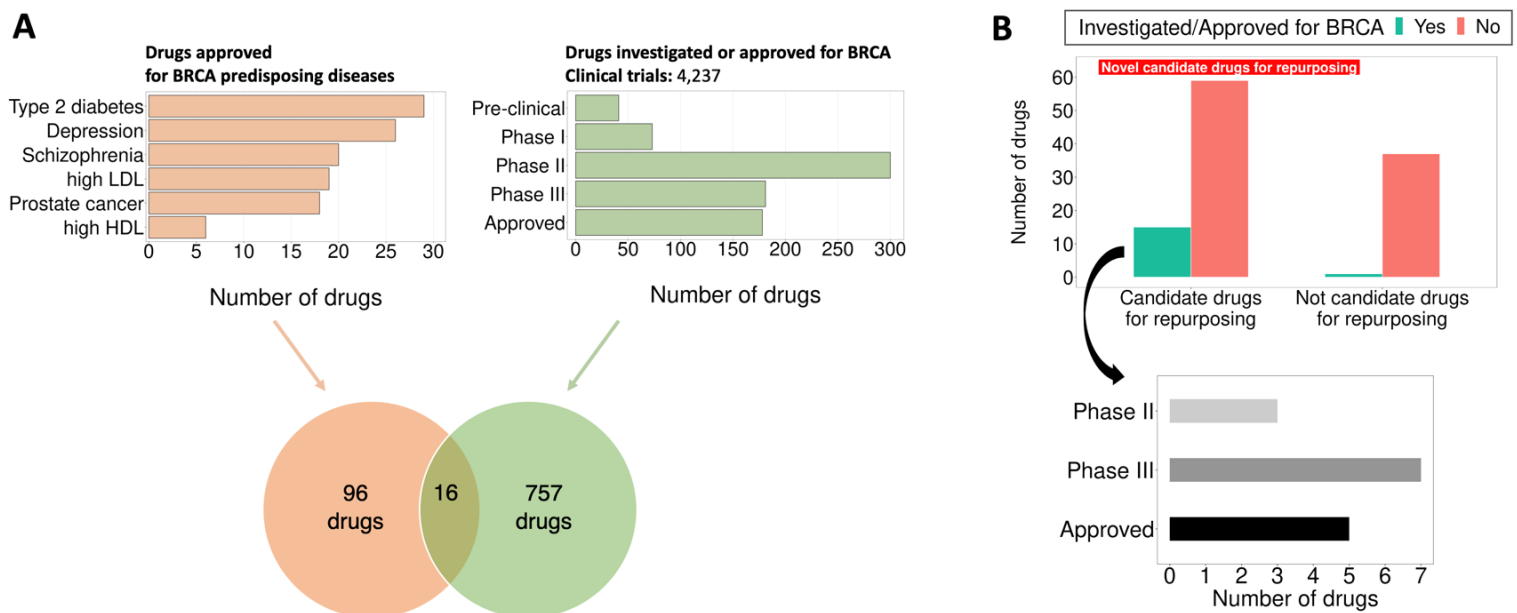


Figure 4. Evaluation of recommended candidate drugs for repurposing in breast cancer. **A.** Out of all drugs approved for breast cancer predisposing diseases, sixteen have been investigated or approved for breast cancer treatment. **B.** Our recommended candidate drugs for repurposing in breast cancer treatment are significantly enriched for drugs already investigated or approved for breast cancer (OR=9.28, $p=7.99e-03$, one-sided Fisher's exact test). Notably, our list of candidate drugs includes 15 of the 16 drugs currently in advanced clinical trials (Phase II or III) or approved for breast cancer treatment. BRCA: breast cancer

Discussion

In our previous work, we exploited monogenic diseases that predispose their bearers to complex diseases, showing that drugs targeting the causal gene of the monogenic disease are good candidates for treating the associated complex disease²⁰. With that study we introduced the idea that predisposing health conditions of a complex disease, and their genetics, can be used to inform drug repurposing for the complex disease. Building on that

work, here, we develop an approach to investigate whether complex disease risk factors that predispose individuals to a complex disease can inform the repurposing of existing drugs for the complex disease. To test this, we use breast cancer as an example of a well-studied and common complex disease, with many predisposing risk factors. We show that shared genes and pathways between breast cancer and its predisposing diseases can point to promising candidate drugs for repurposing for breast cancer treatment that target the shared pathways. Importantly, our recommended candidate drugs are enriched for those currently undergoing clinical trials or already approved for breast cancer treatment, highlighting the ability of our approach in identifying likely successful candidate drugs for repurposing.

Pleiotropy has been previously used for the identification of shared biology between diseases from the same body systems, such as psychiatric diseases. However, our study aims to identify shared genetic factors between breast cancer and its predisposing diseases which primarily affect different body systems, such as depression and high HDL. Despite affecting distinct anatomical regions, these disease pairs share pathophysiological processes which suggests that association is not just due to the known similarity between the phenotypes. Our approach identifies these pleiotropic genes using a local genetic correlation analysis and links them to canonical pathways which are used to prioritize candidate drugs for repurposing. With that, our study is the first to propose an alternative way to analyze publicly available GWAS summary data that can both suggest new uses for existing drugs and point towards the genes and pathways supporting drug repurposing.

We present an example to illustrate the power of our approach in both identifying shared biology between diseases and providing a biological basis for the repurposing of a drug. It is known that elevated levels of HDL increase the risk for breast cancer²². By analyzing this pair of diseases, we identify *MLXIPL* as a shared gene between breast cancer and high HDL. We also find that *MLXIPL* is significantly connected to the *FOXA2* pathway. Interestingly, the *FOXA2* pathway is known to play a role in both breast cancer pathogenesis and lipid metabolism (pleiotropic effect)^{38–40}. Among the approved lipid-lowering medications, we find two fibric acids (fenofibrate and gemfibrozil) and one statin (rosuvastatin) significantly connected to the *FOXA2* pathway. Notably, both drug categories have demonstrated anti-cancer properties in preclinical and clinical studies, respectively^{41,42}. This example shows that our approach can identify meaningful biological signals. It also shows that by leveraging the shared biology between breast cancer and its predisposing factors, we can prioritize candidate drugs for repurposing while also providing plausible biological mechanisms through which these drugs may impact breast cancer.

Our approach has some limitations. First, we define shared genes as those located within the detected, positively correlated shared genomic loci for each disease pair. However, a SNP in a shared locus might have distal regulatory effects on genes located outside that locus and including those genes could potentially result in a more complete list of shared genes for each disease pair. Second, we rely on pleiotropy to identify shared genetic factors between breast cancer and its predisposing diseases. However, it is possible that a disease might increase the risk for breast cancer through indirect or interaction effects, which are not captured by our approach. Third, redundant canonical pathways in the Molecular Signatures Database (MSigDB) could result in the identification of fewer significantly shared pathways between diseases than their actual number due to multiple testing corrections. Fourth, we analyzed GWAS summary statistics data only from the European population, due to greater sample sizes and data availability. Future studies analyzing data from diverse populations may discover shared loci missing from our analysis.

In conclusion, our approach detects and leverages the shared biology between pairs of breast cancer with its predisposing diseases to suggest novel drugs for breast cancer treatment, while also providing a biological basis for each drug recommendation. Future work can exploit our list of candidate drugs for repurposing and evaluate their efficacy in experimental settings. While we specifically applied our approach to the case of breast cancer, we trust that it can be applied to recommend novel candidate drugs for repurposing for any complex disease with a known set of predisposing diseases and available GWAS summary statistics. Therefore, it serves as a valuable tool for not only identifying promising candidate drugs for repurposing across various complex diseases, but also suggesting the disease networks involved in repurposing.

Methods

Data collection and preprocessing

Genetics data

We download the most recent, publicly available GWAS summary statistics data of European ancestry for breast cancer⁴³, depression⁴⁴, high HDL⁴⁵, high LDL⁴⁵, prostate cancer⁴⁶, schizophrenia⁴⁷ and type 2 diabetes⁴⁸. For each GWAS, we keep only bi-allelic SNPs in autosomal chromosomes that match to unique rsIDs. We also harmonize the effect and alternate allele for each SNP using the 1K Genomes reference panel of Europeans. The GWAS for schizophrenia included an imputation quality score for each SNP and we filter for SNPs with imputation score ≥ 0.3 .

Protein-protein interaction network

We download the STRING protein-protein interaction (PPI) network for humans (version 11.5, date of download: May 8, 2023). This network contains both physical and functional interactions of proteins and assigns a confidence score to each interaction (edge) that represents the amount of evidence supporting it. We use this score to filter for edges with high confidence (score ≥ 0.7). We also remove multiple edges and loops and map each gene-node to entrezIDs using a file provided by STRING ("9606.protein.aliases.v11.5.txt"). The final PPI network contains 16,115 nodes and 240,541 edges.

Canonical pathways

We download a list of gene sets for canonical pathways from the MSigDB (version 2023.2). This is a list of manually curated gene sets by domain experts and includes information from five databases: BioCarta, KEGG (Kyoto Encyclopedia of Genes and Genomes), Reactome, PID (Pathway Interaction Database) and WikiPathways. For each canonical pathway, we filter for genes found in the STRING PPI network. Eventually, we have a list of 3,795 canonical pathways that we use to connect shared genes to drugs.

Drugs indicated for breast cancer predisposing diseases

For each of the six breast cancer predisposing diseases (depression, high HDL, high LDL, prostate cancer, schizophrenia, type 2 diabetes), we compile a list of FDA-approved drugs using RxNORM and annotate them with gene-targets from DrugBank (version 5.19, date of

download: June 16, 2022). We do not filter the list of targets based on pharmacological action. Finally, we keep the drug gene-target with unique entrezIDs that are found in the STRING PPI network.

Drugs investigated or indicated for breast cancer

To find drugs currently investigated for breast cancer, we download clinical trial data from the Aggregate Content of ClinicalTrials.gov (AACT) database in a pipe-delimited format (date of download: November 4, 2022; note that it is updated daily). AACT is a publicly available relational database that contains information about all the studies registered in ClinicalTrials.gov. We obtain information for 432,597 clinical trials that were registered in ClinicalTrials.gov by the date of download. Using a manually curated list of MeSH codes related to breast cancer (D001943, D000071960, D018270, D018275, D061325, D058922, D064726, D000069584), we keep 4,237 clinical trials that tested 774 drugs for breast cancer treatment.

To find drugs currently indicated for breast cancer, we obtain the indications of all FDA-approved drugs from RxNORM and filter for those that treat breast cancer using the MeSH terms mentioned above. We find 68 breast cancer indicated drugs and combine them with the investigated drugs to get a complete list of drugs for breast cancer. Again, all drugs are coded in DrugBank IDs.

Finding likely shared genes between breast cancer and its predisposing diseases

To identify shared genes between each pair of breast cancer and predisposing diseases, we use GWAS summary statistics data and LOGODetect (default settings) to perform a local genetic correlation analysis. LOGODetect scans the entire genome to identify loci with correlated SNP effects in two GWAS, accounting for linkage disequilibrium. Following the identification of shared loci for each disease pair, we use the FUMA SNP2GENE function to compile a list of protein-coding genes positionally located within and ± 10 kb upstream and downstream of positively correlated loci (default setting). We do this by providing to FUMA SNP2GENE a list of all SNPs within the identified loci and not only the genome-wide significant ones. Eventually, for each breast cancer and predisposing disease pair, we obtain a list of protein-coding genes (entrezIDs) that are part of their shared etiology (termed as shared genes).

We then seek to prioritize genes that are more likely to contribute to the shared biology for each disease pair. To achieve this, we utilize two commonly used tools: MAGMA (gene-based) and S-MultiXcan (cis-eQTL based). First, we run MAGMA with default settings to obtain gene-based p-values for each disease and adjust them for multiple hypothesis testing using the Benjamini-Hochberg correction method. Second, we use S-MultiXcan to obtain the genetically predicted expression of genes, summarized across 43 GTEx tissues. Then, for each disease pair, we filter the previously identified shared genes for those that are significantly associated with the predisposing disease (MAGMA $p_{\text{adjusted}} < 0.05$) and are predicted to be dysregulated in the same direction in both diseases (sign of S-MultiXcan calculated z-score). Ultimately, these two filtering steps, one gene-based and one cis-eQTL-based, help us prioritize shared genes more likely to contribute to the shared biology of a breast cancer and predisposing disease pair.

Connecting shared genes to drugs via canonical pathways

Our goal here is to link the shared genes with drugs and gain insights into the implicated biological pathways. To do so, we first find which canonical pathways are significantly connected to the shared genes using the STRING PPI network and the Personalized PageRank (PPR) algorithm. PPR uses a user-defined gene, named seed gene, as a starting point and performs random walks in the PPI network, with a probability (default=0.8) of returning to the seed gene after each step. This process assigns a score to each gene based on its connectivity with the seed gene: genes with higher connectivity receive higher scores. Using each shared gene for a disease pair as a seed gene, we calculate the average PPR score of genes within a canonical pathway. To determine if the observed score is higher than expected by chance, we compare it to random PPR scores obtained through 1,000 permutations (more details in “Statistical analysis - Permutation tests”). Ultimately, for each breast cancer and predisposing disease pair, this process yields a list of canonical pathways that are significantly connected ($p_{\text{permutation}} < 0.05$) to the previously identified shared genes (referred to as shared canonical pathways).

We then seek to link the shared canonical pathways to drugs approved for a predisposing disease. To do so, we repeat the above analysis but this time, we use a target of a drug as a seed gene (instead of a shared gene). Similarly, we calculate the average PPR score of genes within a shared canonical pathway and compare it to average PPR scores obtained after 1,000 permutations. Drugs approved for a predisposing disease and significantly connected to a shared canonical pathway ($p_{\text{permutation}} < 0.05$), are considered to target the shared etiology with breast cancer. Therefore, these drugs constitute the prioritized candidate drugs for repurposing for breast cancer, and the shared genes and canonical pathways they target provide biological insights for their effect on breast cancer.

Statistical analysis

Evaluating candidate drugs for repurposing

We find 76 unique drugs approved for a predisposing disease that can be linked with its shared biology with breast cancer. Using these drugs alongside those currently in clinical trials or approved for breast cancer treatment, we test for significant overlap using the Fisher's one-sided exact test.

Permutation tests

To assess the significance of the observed associations between shared genes, drugs, and canonical pathways, we conduct permutations tests.

First, for each breast cancer and predisposing disease pair, we connect each canonical pathway to the identified shared genes. To do so, we use each shared gene as a seed gene, and we calculate the average PPR score for the genes in a canonical pathway. To determine the significance of these connections, we create a null distribution of average PPR scores for genes within a canonical pathway. This involves grouping all genes in the PPI network into four bins based on their degree (quantiles). Subsequently, for each seed (shared) gene, we randomly select 1,000 genes from the same degree bin and calculate the average PPR score for genes within the canonical pathway. These scores are then compared to the observed average PPR score, and a permuted p-value is calculated based on how often the observed score is lower than the permuted scores. This process yields a permuted p-value for every

canonical pathway-shared gene pair, which is then adjusted for the number of shared genes and canonical pathways tested using the Bonferroni correction method.

Second, we link the shared canonical pathways to drugs approved for a predisposing disease to prioritize candidate drugs for repurposing. To do so, we follow a similar approach as above, but this time, we use random genes matched to the gene-target of each drug using the four degree bins. This process yields a permuted p-value for every shared canonical pathway-drug target pair, which is then adjusted for the number of drug targets tested using the Bonferroni correction method.

References

1. Guo Q, Lu Y, Liu W, Lan G, Lan T. The global, regional, and national disease burden of breast cancer attributable to tobacco from 1990 to 2019: a global burden of disease study. *BMC Public Health*. 2024;24(1):107. doi:10.1186/s12889-023-17405-w
2. Waks AG, Winer EP. Breast Cancer Treatment: A Review. *JAMA*. 2019;321(3):288-300. doi:10.1001/jama.2018.19323
3. Franzoi MA, Agostinetti E, Perachino M, et al. Evidence-based approaches for the management of side-effects of adjuvant endocrine therapy in patients with breast cancer. *Lancet Oncol*. 2021;22(7):e303-e313. doi:10.1016/S1470-2045(20)30666-5
4. Ferreira AR, Di Meglio A, Pistilli B, et al. Differential impact of endocrine therapy and chemotherapy on quality of life of breast cancer survivors: a prospective patient-reported outcomes analysis. *Ann Oncol Off J Eur Soc Med Oncol*. 2019;30(11):1784-1795. doi:10.1093/annonc/mdz298
5. Aggarwal S, Verma SS, Aggarwal S, Gupta SC. Drug repurposing for breast cancer therapy: Old weapon for new battle. *Semin Cancer Biol*. 2021;68:8-20. doi:10.1016/j.semcancer.2019.09.012
6. Lord SR, Harris AL. Is it still worth pursuing the repurposing of metformin as a cancer therapeutic? *Br J Cancer*. 2023;128(6):958-966. doi:10.1038/s41416-023-02204-2
7. De A, Kuppusamy G. Metformin in breast cancer: preclinical and clinical evidence. *Curr Probl Cancer*. 2020;44(1):100488. doi:10.1016/j.currprobcancer.2019.06.003
8. Schlander M, Hernandez-Villafuerte K, Cheng CY, Mestre-Ferrandiz J, Baumann M. How Much Does It Cost to Research and Develop a New Drug? A Systematic Review and Assessment. *PharmacoEconomics*. 2021;39(11):1243-1269. doi:10.1007/s40273-021-01065-y
9. Wong CH, Siah KW, Lo AW. Estimation of clinical trial success rates and related parameters. *Biostatistics*. 2019;20(2):273-286. doi:10.1093/biostatistics/kxx069
10. Dowden H, Munro J. Trends in clinical success rates and therapeutic focus. *Nat Rev Drug Discov*. 2019;18(7):495-496. doi:10.1038/d41573-019-00074-z
11. Nelson MR, Tipney H, Painter JL, et al. The support of human genetic evidence for approved drug indications. *Nat Genet*. 2015;47(8):856-860. doi:10.1038/ng.3314
12. King EA, Davis JW, Degner JF. Are drug targets with genetic support twice as likely to be approved? Revised estimates of the impact of genetic support for drug mechanisms on the probability of drug approval. Marchini J, ed. *PLOS Genet*. 2019;15(12):e1008489. doi:10.1371/journal.pgen.1008489
13. Minikel EV, Painter JL, Dong CC, Nelson MR. Refining the impact of genetic evidence on clinical success. *Nature*. Published online April 17, 2024. doi:10.1038/s41586-024-07316-0

14. NBCS Collaborators, ABCTB Investigators, ConFab/AOCS Investigators, et al. Association analysis identifies 65 new breast cancer risk loci. *Nature*. 2017;551(7678):92-94. doi:10.1038/nature24284
15. Reay WR, Cairns MJ. Advancing the use of genome-wide association studies for drug repurposing. *Nat Rev Genet*. 2021;22(10):658-671. doi:10.1038/s41576-021-00387-z
16. Tam V, Patel N, Turcotte M, Bossé Y, Paré G, Meyre D. Benefits and limitations of genome-wide association studies. *Nat Rev Genet*. 2019;20(8):467-484. doi:10.1038/s41576-019-0127-1
17. Schaid DJ, Chen W, Larson NB. From genome-wide associations to candidate causal variants by statistical fine-mapping. *Nat Rev Genet*. 2018;19(8):491-504. doi:10.1038/s41576-018-0016-z
18. Nam Y, Jung SH, Yun JS, et al. Discovering comorbid diseases using an inter-disease interactivity network based on biobank-scale PheWAS data. *Bioinforma Oxf Engl*. 2023;39(1):btac822. doi:10.1093/bioinformatics/btac822
19. Melamed RD, Emmett KJ, Madubata C, Rzhetsky A, Rabadan R. Genetic similarity between cancers and comorbid Mendelian diseases identifies candidate driver genes. *Nat Commun*. 2015;6(1):7033. doi:10.1038/ncomms8033
20. Lalagkas PN, Melamed RD. Shared etiology of Mendelian and complex disease supports drug discovery. Published online April 19, 2024. doi:10.21203/rs.3.rs-4250176/v1
21. Zhu GL, Xu C, Yang K bin, et al. Causal relationship between genetically predicted depression and cancer risk: a two-sample bi-directional mendelian randomization. *BMC Cancer*. 2022;22(1):353. doi:10.1186/s12885-022-09457-9
22. Johnson KE, Siewert KM, Klarin D, et al. The relationship between circulating lipids and breast cancer risk: A Mendelian randomization study. Minelli C, ed. *PLOS Med*. 2020;17(9):e1003302. doi:10.1371/journal.pmed.1003302
23. Ren ZJ, Cao DH, Zhang Q, et al. First-degree family history of breast cancer is associated with prostate cancer risk: a systematic review and meta-analysis. *BMC Cancer*. 2019;19(1):871. doi:10.1186/s12885-019-6055-9
24. Cui Y, Lu W, Shao T, Zhuo Z, Wang Y, Zhang W. Severe mental illness and the risk of breast cancer: A two-sample, two-step multivariable Mendelian randomization study. Montazeri A, ed. *PLOS ONE*. 2023;18(9):e0291006. doi:10.1371/journal.pone.0291006
25. Pearson-Stuttard J, Papadimitriou N, Markozannes G, et al. Type 2 Diabetes and Cancer: An Umbrella Review of Observational and Mendelian Randomization Studies. *Cancer Epidemiol Biomarkers Prev*. 2021;30(6):1218-1228. doi:10.1158/1055-9965.EPI-20-1245
26. Guo H, Li JJ, Lu Q, Hou L. Detecting local genetic correlations with scan statistics. *Nat Commun*. 2021;12(1):2033. doi:10.1038/s41467-021-22334-6
27. Watanabe K, Taskesen E, Van Bochoven A, Posthuma D. Functional mapping and annotation of genetic associations with FUMA. *Nat Commun*. 2017;8(1):1826. doi:10.1038/s41467-017-01261-5
28. De Leeuw CA, Mooij JM, Heskes T, Posthuma D. MAGMA: Generalized Gene-Set Analysis of GWAS Data. Tang H, ed. *PLOS Comput Biol*. 2015;11(4):e1004219. doi:10.1371/journal.pcbi.1004219
29. Barbeira AN, Pividori M, Zheng J, Wheeler HE, Nicolae DL, Im HK. Integrating predicted transcriptome from multiple tissues improves association detection. Plagnol V, ed. *PLOS Genet*. 2019;15(1):e1007889. doi:10.1371/journal.pgen.1007889
30. Ghousaini M, Nelson MR, Dunham I. Future prospects for human genetics and genomics in drug discovery. *Curr Opin Struct Biol*. 2023;80:102568. doi:10.1016/j.sbi.2023.102568

31. Xu X, Zhang M, Xu F, Jiang S. Wnt signaling in breast cancer: biological mechanisms, challenges and opportunities. *Mol Cancer*. 2020;19(1):165. doi:10.1186/s12943-020-01276-5
32. Sheng R, Kim H, Lee H, et al. Cholesterol selectively activates canonical Wnt signalling over non-canonical Wnt signalling. *Nat Commun*. 2014;5(1):4393. doi:10.1038/ncomms5393
33. Murillo-Garzón V, Kypta R. WNT signalling in prostate cancer. *Nat Rev Urol*. 2017;14(11):683-696. doi:10.1038/nrurol.2017.144
34. Chen J, Ning C, Mu J, Li D, Ma Y, Meng X. Role of Wnt signaling pathways in type 2 diabetes mellitus. *Mol Cell Biochem*. 2021;476(5):2219-2232. doi:10.1007/s11010-021-04086-5
35. Ortega MA, Fraile-Martínez O, Asúnsolo Á, Buján J, García-Honduvilla N, Coca S. Signal Transduction Pathways in Breast Cancer: The Important Role of PI3K/Akt/mTOR. *J Oncol*. 2020;2020:1-11. doi:10.1155/2020/9258396
36. Hare SH, Harvey AJ. mTOR function and therapeutic targeting in breast cancer. *Am J Cancer Res*. 2017;7(3):383-404.
37. Laplante M, Sabatini DM. An Emerging Role of mTOR in Lipid Biosynthesis. *Curr Biol*. 2009;19(22):R1046-R1052. doi:10.1016/j.cub.2009.09.058
38. Zhang Z, Yang C, Gao W, et al. FOXA2 attenuates the epithelial to mesenchymal transition by regulating the transcription of E-cadherin and ZEB2 in human breast cancer. *Cancer Lett*. 2015;361(2):240-250. doi:10.1016/j.canlet.2015.03.008
39. Perez-Balaguer A, Ortiz-Martínez F, García-Martínez A, Pomares-Navarro C, Lerma E, Peiró G. FOXA2 mRNA expression is associated with relapse in patients with Triple-Negative/Basal-like breast carcinoma. *Breast Cancer Res Treat*. 2015;153(2):465-474. doi:10.1007/s10549-015-3553-6
40. Wolfrum C, Asilmaz E, Luca E, Friedman JM, Stoffel M. Foxa2 regulates lipid metabolism and ketogenesis in the liver during fasting and in diabetes. *Nature*. 2004;432(7020):1027-1032. doi:10.1038/nature03047
41. Li T, Zhang Q, Zhang J, et al. Fenofibrate induces apoptosis of triple-negative breast cancer cells via activation of NF- κ B pathway. *BMC Cancer*. 2014;14(1):96. doi:10.1186/1471-2407-14-96
42. Jiang W, Hu JW, He XR, Jin WL, He XY. Statins: a repurposed drug to fight cancer. *J Exp Clin Cancer Res*. 2021;40(1):241. doi:10.1186/s13046-021-02041-2
43. Zhang H, Ahearn TU, Lecarpentier J, et al. Genome-wide association study identifies 32 novel breast cancer susceptibility loci from overall and subtype-specific analyses. *Nat Genet*. 2020;52(6):572-581. doi:10.1038/s41588-020-0609-2
44. Howard DM, Adams MJ, Clarke TK, et al. Genome-wide meta-analysis of depression identifies 102 independent variants and highlights the importance of the prefrontal brain regions. *Nat Neurosci*. 2019;22(3):343-352. doi:10.1038/s41593-018-0326-7
45. Graham SE, Clarke SL, Wu KHH, et al. The power of genetic diversity in genome-wide association studies of lipids. *Nature*. 2021;600(7890):675-679. doi:10.1038/s41586-021-04064-3
46. The Profile Study, Australian Prostate Cancer BioResource (APCB), The IMPACT Study, et al. Association analyses of more than 140,000 men identify 63 new prostate cancer susceptibility loci. *Nat Genet*. 2018;50(7):928-936. doi:10.1038/s41588-018-0142-8
47. Trubetskoy V, Pardiñas AF, Qi T, et al. Mapping genomic loci implicates genes and synaptic biology in schizophrenia. *Nature*. 2022;604(7906):502-508. doi:10.1038/s41586-022-04434-5

48. Mahajan A, Taliun D, Thurner M, et al. Fine-mapping type 2 diabetes loci to single-variant resolution using high-density imputation and islet-specific epigenome maps. *Nat Genet.* 2018;50(11):1505-1513. doi:10.1038/s41588-018-0241-6